# Data Quality Challenges in Distributed Live-Virtual-Constructive Test Environments

JEREMY R. MILLAR, DOUGLAS D. HODSON, GILBERT L. PETERSON and DARRYL K. AHNER, Air Force Institute of Technology

## 1. INTRODUCTION

Live-virtual-constructive (LVC) simulations are complex systems comprising a combination of live (real people operating real equipment), virtual (real people operating simulated equipment, or vice versa), and constructive (wholly simulated) entities. Nodes in the system support the simulation of one or more entities and are often geographically distributed to leverage unique assets, e.g., physical test range space or high-fidelity full motion simulators. Nodes are connected in a peer-to-peer fashion and communicate using protocols such as Distributed Interactive Simulation (DIS) [DIS Steering Committee 1998], the High Level Architecture (HLA) [Dahmann et al. 1997], or the Test and Training Enabling Network Architecture (TENA) [Powell and Noseworthy 2012].

Distributed LVC simulation promises a number of benefits for the test and evaluation (T&E) community, including reduced costs, access to simulations of limited availability assets, the ability to conduct large-scale multi-service test events, and recapitalization of existing simulation investments. Consequently, the Department of Defense (DoD) is increasingly turning to LVC simulation and virtual environments to support T&E events. LVC simulations have been used to test communications for unmanned aircraft systems [Parker et al. 2009], conduct cyber security analysis [Van Leeuwen et al. 2010], and quantify radar measurement errors [Hodson et al. 2013].

Ensuring rigorous results for T&E events supported by LVC simulation requires addressing three fundamental data quality challenges: quantifying numerical errors due to weakly consistent nodes; assessing measurement accuracy with respect to tolerance requirements; and assessing measurment quality in the absence of absolute truth values.

## 2. DATA QUALITY CHALLENGES

LVC simulations are typically designed as fully replicated, geographically distributed database applications with real-time constraints. The data of interest is composed of entity and world state information and derived quantities such as collisions or weapons effectiveness. This data must be replicated at each node to meet availability and responsiveness requirements. The inclusion of live or virtual entities imposes real-time constraints on database responsiveness since long read/write latencies cannot be tolerated. Consequently, entity state updates (i.e., writes to a database record) are propagated to other system nodes after taking effect locally and are delayed due to network latency. Thus, not all nodes see the same simulation state at the same time. If updates cease, the system will eventually become consistent [Terry et al. 1994]. As such, LVC simulations can be viewed in the same context as eventually consistent distributed datastores such as Amazon's Dynamo [DeCandia et al. 2007], Cassandra [Lakshman and Malik 2009], or Megastore [Baker et al. 2011].

For any eventually consistent distributed database, a fundamental question is "How eventual is eventual?" Common measures of eventual consistency are time (how long does it take for readers to see the result of a write) and versions (how many versions old is a given read result) [Bailis and Ghodsi 2013]. For LVC simulations and other distributed virtual environments, deviation (e.g., Euclidean distance) from a "true" value is a common measure of consistency [Yu and Vahdat 2002; Aggarwal et al. 2004; Zhou et al. 2004].

Quantifying the numerical error associated with eventual consistency is a key challenge for LVC simulations. While there is a growing body of literature characterizing the consistency of distributed databases such as Dynamo and Cassandra [Wada et al. 2011; Rahman et al. 2012; Bailis et al. 2012], these works focus on time or version staleness as the measure of consistency, feature read-heavy workloads, and are not necessarily geographically distributed. In contrast, LVC simulations are more concerned with numerical error, have a balanced read/write workload, and are distributed geographically.

During a test event, measurements may be taken at any simulation node. Inconsistencies in the replicated state are reflected as measurement errors. A second challenge lies in assessing whether each measurement error lies within a precision tolerance. This assessment must be conducted during system design to ensure the simulation is capable of meeting test requirements. Additionally, it must occur during the test execution to provide a quantification of the uncertainty associated with each measurement.

A third challenge for LVC simulations is assessing the quality of measurements without a known truth value, particularly in the discrete case [Mauve 2000]. This is especially true for derived quantities that depend on inconsistent state data such as collisions and weapons effects. In this case, each interacting node may compute a result that is correct according to its state replica and different from other interacting nodes. Furthermore, the uncertainty can vary based on the node taking the measurement.

## 3. CONCLUSION

LVC simulations enable large-scale operationally relevant T&E events at reduced cost, provide access to limited availability assets, and recapitalize existing simulations. System architectures based on weakly consistent replicated databases and unreliable update protocols yield three fundamental challenges for data quality: quantifying error due to eventual consistency; assessing measurment accuracy with respect to desired tolerances; and assessing measurement quality in the absence of a truth value. Addressing these challenges is fundamental to ensuring the veracity and rigor of T&E events supported by LVC simulation.

## REFERENCES

Sudhir Aggarwal, Hemant Banavar, Amit Khandelwal, Sarit Mukherjee, and Sampath Rangarajan. 2004. Accuracy in dead-reckoning based distributed multi-player games. In *Proceedings of 3rd ACM SIG-COMM workshop on Network and system support for games*. ACM, 161–165.

Peter Bailis and Ali Ghodsi. 2013. Eventual consistency today: limitations, extensions, and beyond. *Commun. ACM* 56, 5 (May 2013), 55–63. DOI:http://dx.doi.org/10.1145/2447976.2447992

Peter Bailis, Shivaram Venkataraman, Michael J. Franklin, Joseph M. Hellerstein, and Ion Stoica. 2012. Probabilistically bounded staleness for practical partial quorums. *Proc. VLDB Endow.* 5, 8 (April 2012), 776–787. http://dl.acm.org/citation.cfm?id=2212351.2212359

Jason Baker, Chris Bond, James C Corbett, JJ Furman, Andrey Khorlin, James Larson, Jean-Michel Leon, Yawei Li, Alexander Lloyd, and Vadim Yushprakh. 2011. Megastore: Providing Scalable, Highly Available Storage for Interactive Services.. In *CIDR*, Vol. 11. 223–234.

Judith S Dahmann, Richard M Fujimoto, and Richard M Weatherly. 1997. The department of defense high level architecture. In *Proceedings of the 29th conference on Winter simulation*. IEEE Computer Society, 142–149.

Giuseppe DeCandia, Deniz Hastorun, Madan Jampani, Gunavardhan Kakulapati, Avinash Lakshman, Alex Pilchin, Swaminathan Sivasubramanian, Peter Vosshall, and Werner Vogels. 2007. Dynamo: amazon's highly available key-value store. In *ACM SIGOPS Operating Systems Review*, Vol. 41. ACM, 205–220.

DIS Steering Committee. 1998. IEEE Standard for Distributed Interactive Simulation-Application Protocols. *IEEE Standard* 1278 (1998).

Douglas D Hodson, Alex J Gutman, Bruce Esken, and Raymond R Hill. 2013. Quantifying Radar Measurement Errors in a Live-Virtual-Constructive Environment to Determine System Viability: A Case Study. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* (2013), 1548512913503740.

Avinash Lakshman and Prashant Malik. 2009. Cassandra: structured storage system on a p2p network. In *Proceedings of the 28th ACM symposium on Principles of distributed computing*. ACM, 5–5.

Martin Mauve. 2000. How to keep a dead man from shooting. In *Interactive Distributed Multimedia Systems and Telecommunication Services*. Springer, 199–204.

Eric Paul Parker, Nadine Elizabeth Miner, Brian Peter Van Leeuwen, and James Brian Rigdon. 2009. Testing unmanned autonomous system communications in a Live/Virtual/Constructive environment. *International Test and Evaluation Association Journal (ITEA)* 30 (2009), 513–522.

Edward T Powell and J Russell Noseworthy. 2012. The test and training enabling architecture (TENA). *Engineering Principles of Combat Modeling and Distributed Simulation* (2012), 449.

Muntasir Raihan Rahman, Wojciech Golab, Alvin AuYoung, Kimberly Keeton, and Jay J Wylie. 2012. Toward a principled framework for benchmarking consistency. In *Proceedings of the Eighth USENIX conference on Hot Topics in System Dependability*. USENIX Association, 8–8.

Douglas B Terry, Alan J Demers, Karin Petersen, Mike J Spreitzer, Marvin M Theimer, and Brent B Welch. 1994. Session guarantees for weakly consistent replicated data. In *Parallel and Distributed Information Systems, 1994., Proceedings of the Third International Conference on*. IEEE, 140–149.

Brian Van Leeuwen, Vincent Urias, John Eldridge, Charles Villamarin, and Ron Olsberg. 2010. Performing cyber security analysis using a live, virtual, and constructive (LVC) testbed. In *Military Communications Conference, 2010-MILCOM 2010*. IEEE, 1806–1811.

Hiroshi Wada, Alan Fekete, Liang Zhao, Kevin Lee, and Anna Liu. 2011. Data Consistency Properties and the Trade-offs in Commercial Cloud Storage: the Consumers' Perspective.. In *CIDR*, Vol. 11. 134–143.

Haifeng Yu and Amin Vahdat. 2002. Design and evaluation of a conit-based continuous consistency model for replicated services. *ACM Transactions on Computer Systems (TOCS)* 20, 3 (2002), 239–282.

Suiping Zhou, Wentong Cai, Bu-Sung Lee, and Stephen J Turner. 2004. Time-space consistency in large-scale distributed virtual environments. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 14, 1 (2004), 31–47.